How Long Will it Take:
A Historical and Projective Approach to Boundary Crossing

Victor de la Peña, Columbia University

Joint work with Brown, M.[#], Kushnir, Y.[*], Ravindranath, A.[†] & Sit, T.[†]

[#]Department of Mathematics, City University of New York, New York, NY
[†] Department of Statistics, Columbia University, New York, NY
[*] Lamont-Doherty Earth Observatory, The Earth Institute at Columbia University, Palisades, NY

February 2011

Outline

Part I: An application of Boundary Crossing in Climatology

1. Introduction
2. Methodology
3. Examples and results

Part II: From Boundary Crossing of Non-random Functions to Boundary crossing of stochastic Processes

1. Bounding the first passage time on an average

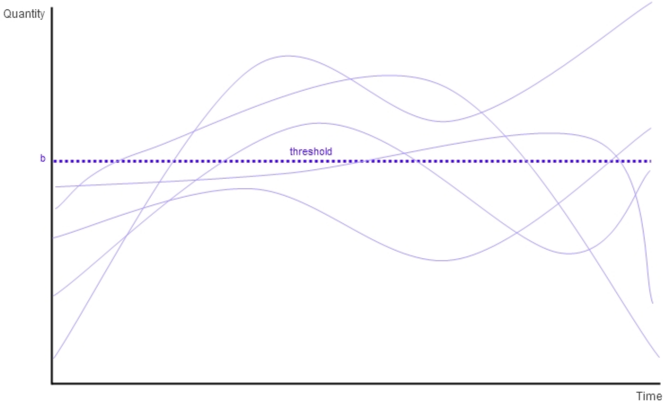
Part I: An application of Boundary Crossing in Climatology

Introduction

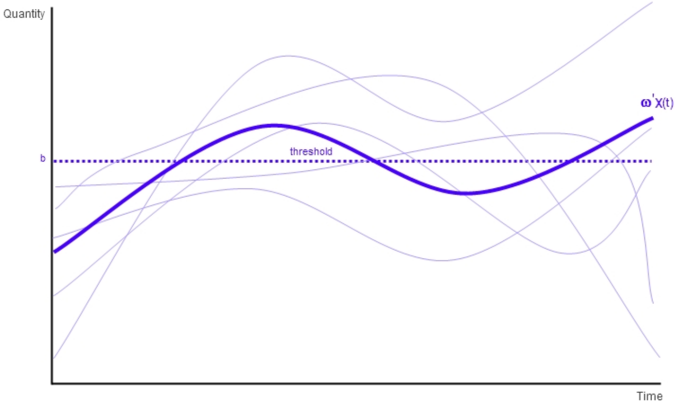
- The study of the projected path of the climate system involves an assessment of the crossing of significant thresholds referred to as "impact threshold" or "tipping points".

- "Impact threshold" refers to *"any degree of change that can link the onset of a given ciritcal biophysical or socio-economeic impact to a particular climate state(s)"* (Pittock & Jones, 2000; Jones, 2001)

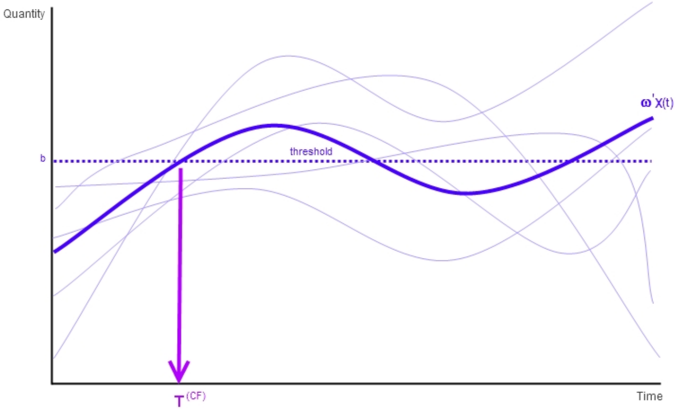
Classical Approach



The Mean Path



Traditional Estimator



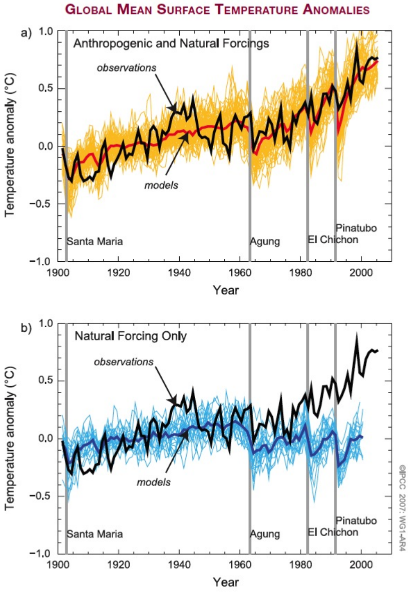
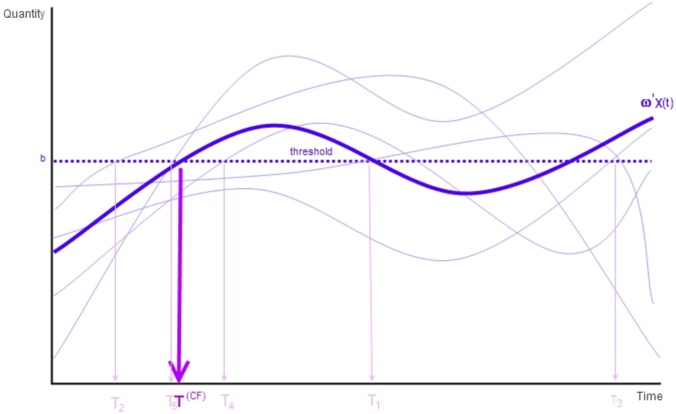


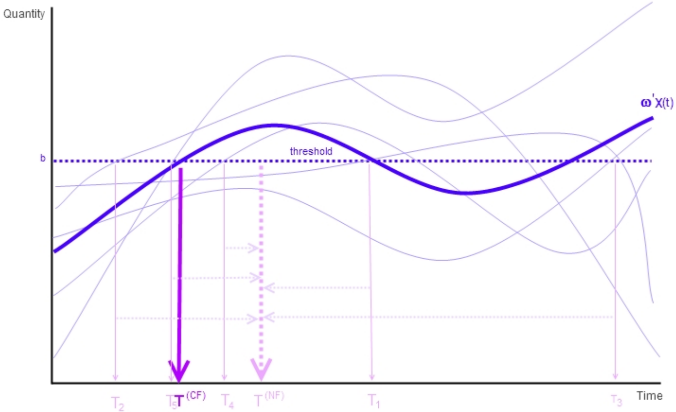
Figure TS.23. *(a) Global mean surface temperature anomalies relative to the period 1901 to 1950, as observed (black line) and as obtained from simulations with both anthropogenic and natural forcings. The thick red curve shows the multi-model ensemble mean and the thin yellow curves show the individual simulations. Vertical grey lines indicate the timing of major volcanic events. (b) As in (a), except that the simulated global mean temperature anomalies are for natural forcings only. The thick blue curve shows the multi-model ensemble mean and the thin lighter blue curves show individual simulations. Each simulation was sampled so that coverage corresponds to that of the observations. {Figure 9.5}*

Adapted from IPCC. (2007).
Climate Change 2007: The Physical Science Basis: Cambridge

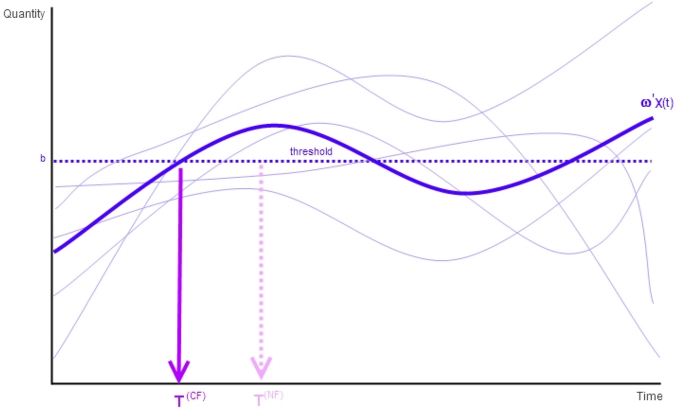
Can we do a better job?



Consider the mean of the stopping times T_i



Now we have two estimators: $T^{(CF)}, T^{(NF)}$



The question that we ask

- In this project, we are concerned with determining the best estimate of a threshold corssing time from a range of climate change projections.

Methodology

- The data used to carry out the demonstration are time series of two subtropical regions:

  (a) the US southwest (125°W to 95°E; 25°N to 40°N) and
  (b) the Mediterranean (10°W to 50°E; 30°N to 45°N)

  calculated from IPCC † Fourth Assessment (AR4) model simulations of the 20th and the 21st century (Randall et al. 2007; Meehl et al. 2007).

† IPCC here refers to Intergovernmental Panel on Climate Change. It is the leading body for the assessment of climate change, established by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) to provide the world with a clear scientific view on the current state of climate change.

Some Technical Details

- Output from 19 models is considered.
- The models are forced in the 20th century with the observed, time dependent greenhouse gas concentrations, anthropogenic aerosols, and volcanic areosols.
- In the future simulations, the models are forced with forcing scenario A1B (see IPCC 2000) - "middle of the road" estimate.

Determining the Threshold

- We use a threshold derived from 19 simulated paths to represent climate states. The models are forced with forcing scenario A1B (see IPCC 2000) - “middle of the road” estimate.
- Specifically, we define the threshold as being one standard deviation below the 21-year averaged rainfall that are sampled annually between 1950 and 2000.

- Assumptions:
  - each of the models provides a/an (exchangeable) realisation of the process under study and
  - the projected paths span the range of the possible future scenarios.

First-hitting Time

- Define

$$T_{r,i} := \inf\{t \in [0,\tau] : X_i(t) \geqslant r\} \quad , i = 1, \ldots, n(= 19),$$

as the first hitting time of the ith simulated path X_i with τ bounded.

- The true path T_b is defined as

$$T_b = \begin{cases} T_1 & \text{with probability (w.p.) } \frac{1}{n}, \\ T_2 & \text{w.p. } \frac{1}{n}, \\ \vdots & \\ T_n & \text{w.p. } \frac{1}{n}. \end{cases}$$

First-hitting Time

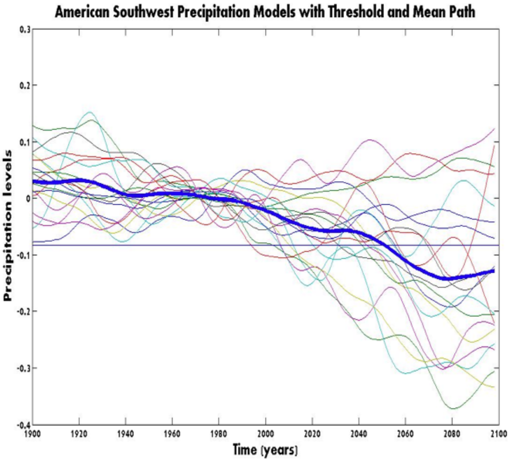
- Two forecasts:
  1. Mean of the first-hitting time:

$$T_r^{(NF)} := \frac{1}{16} \sum_{i=1}^{16} T_{r,i}$$

  2. First-hitting time of the mean path:

$$T_r^{(CF)} := \inf \left\{ t \in [0, \tau] : \frac{1}{16} \sum_{i=1}^{16} X_i(t) \geqslant r \right\}$$

Remark: Only 16 out of the 19 models crossed the threshold before the end of the 21st century. We here exclude the three potential outliers.


American Southwest Precipitation Models with Threshold and Mean Path
Precipitation levels
Time (years)
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
1900
1920
1940
1960
1980
2000
2020
2040
2060
2080
2100

Optimality

[PROPOSITION] Our proposed estimator $T_b^{(NF)}$ outperforms the traditional estimator $T_b^{(CF)}$ in terms of (i) mean-squared error and (ii) Brier skill score.

	$T^{(NF)}_{trunc}$	$T^{(NF)}$	$T^{(CF)}(K^{-1})$
Mediterranean	2010.21	2010.21	2040
Southwest US	2004.63	∞	2018

where $T^{(NF)}_{trunc} = \frac{1}{\sum_{i=1}^{19} \mathbf{1}_{\{T_i<\infty\}}} \sum_{i=1}^{19} T_i \mathbf{1}_{\{T_i<\infty\}}$.

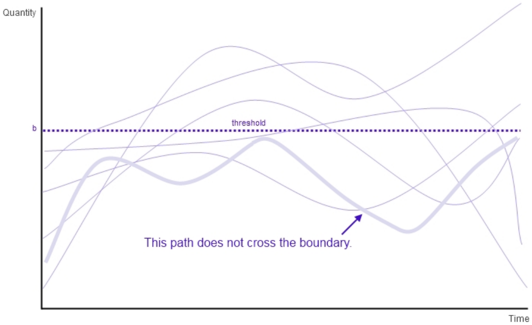
Remarks

- As the figures demonstrate, there is there is discrepancy between the threshold of the average path and the average time of the paths.
- According to current scientific evidence, the transition to a more arid climate in these two regions is already underway, supporting our models' projections.

Part II: From Boundary Crossing of Non-random Functions to Boundary crossing of stochastic Processes

From Boundary Crossing of Non-random Functions to Boundary crossing of stochastic Processes

- What if not all the paths cross the boundary before the end of the experiment?



Intuition

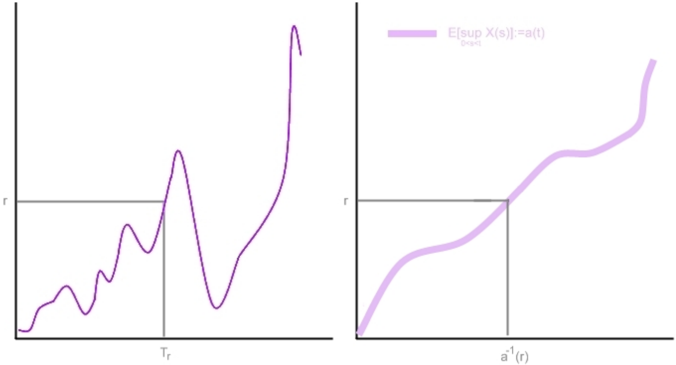


Figure: Extension to Random Processes.

- Question: How is $E[T_r]$ related to $a^{-1}(r)$?

Boundary Crossing: Example

- Let $a_{(n)}(t) = E\sup_{s\leqslant t} X_s = n^{-1}\sum_{i=1}^{n}\sup_{s\leqslant t} Y_{s,i}$. Assume $a_n(t)$ is increasing (we can also use a generalized inverse) with $a_{(n)}^{-1}(r) = t_r = \inf\{t > 0 : a_{(n)}(t) = r\} \longrightarrow a(t)$, we can obtain bounds, under certain conditions:

$$\frac{1}{2}a_{(n)}^{-1}(r/2) \leqslant E[T_r] \leqslant 2a_{(n)}^{-1}(r),$$

when $E[T_r] < \infty$.

Results

	$T^{(NF)}_{trunc}$	$T^{(NF)}$	$T^{(CF)}(K^{-1})$	$\hat{M}$	$a^{-1}_{19}(r)$
Mediterranean	2010.21	2010.21	2040	2018	2008
Southwest US	2004.63	∞	2018	2011	2004

where $T^{(NF)}_{trunc} = \frac{1}{\sum_{i=1}^{19} \mathbf{1}_{\{T_i<\infty\}}} \sum_{i=1}^{19} T_i \mathbf{1}_{\{T_i<\infty\}}$

The End

Proof.

Denote $\mathfrak{F} = \sigma(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(K)})$, where π is a finite permutation, i.e. $\mathfrak{F}$ is the σ-algebra generated by the permutable events of X.

To prove (i) observe that, for any $\mathfrak{F}$-measurable random variable C,

$$E[T_b|\mathfrak{F}] = T^{(NF)} \qquad \text{and}$$

$$\begin{aligned}
S^2_{T_b^{(NF)}, T_b} &:= E\left[T_b^{(NF)} - T_b\right]^2 \\
&= E\left\{E\left[(T_b^{(NF)} - T_b)^2 \middle| \mathfrak{F}\right]\right\} \\
&= E\left\{E\left[(T_b^{(NF)} - C + C - T_b)^2 \middle| \mathfrak{F}\right]\right\} \\
&= E\left(E\left[(T_b^{(NF)} - C)^2 \middle| \mathfrak{F}\right]\right) + 2E\left(E\left[(T_b^{(NF)} - C)(C - T_b) \middle| \mathfrak{F}\right]\right) \\
&\quad + E\left(E\left[(C - T_b)^2 \middle| \mathfrak{F}\right]\right)
\end{aligned}$$

Proof.

$$
\begin{aligned}
&= E\left[T_b^{(NF)} - C\right]^2 + 2E\{(T_b^{(NF)} - C)(C - E\left[T_b^{(NF)} | \mathfrak{F}\right])\} \\
&\quad + E\left(E\left[C - T_b\right]^2 \middle| \mathfrak{F}\right) \\
&= E\left[T_b^{(NF)} - C\right]^2 - 2E\left[T_b^{(NF)} - C\right]^2 + E\left[C - T_b\right]^2 \\
&= E\left[C - T_b\right]^2 - E\left[T_b^{(NF)} - C\right]^2 \\
&\leqslant E\left[C - T_b\right]^2 .
\end{aligned}
$$

The result follows by picking $C = T_b^{(CF)}$. $\square$

Proof.

(ii) follows immediately from (i) by recalling that

$$B_{T_b^{(NF)}, T_b^{(CF)}, T_b} = 1 - \frac{S^2_{T_b^{(NF)}, T_b}}{S^2_{T_b^{(CF)}, T_b}} > B_{T_b^{(CF)}, T_b^{(CF)}, T_b} = 0.$$

$\square$

In fact, the above result holds for any other $\mathcal{F}$-measurable random variable in addition to $T_b^{(CF)}$.

Proof.

Denote $\hat{M}$ the sample median of $\{T_i\}_{i=1}^K$. We are going to prove that $\hat{M}$ minimises the absolute error with respect to T_b. Suppose T_b, the true stopping time, equals the stopping time of one of the simulated paths T_i, with equal probability of K^{-1}, then

$$\begin{aligned} E|T_b - \hat{M}| &= E\{E\left[|T_b - \hat{M}|\middle|\sigma(T_1, \ldots, T_K)\right]\} \\ &= K^{-1}\int \sum_{i=1}^{K} |t_i - \hat{m}| dF(\mathbf{t}) \\ &\leqslant K^{-1}\int \sum_{i=1}^{K} |t_i - c(\mathbf{t})| dF(\mathbf{t}) \\ &= E|T_b - C|, \end{aligned}$$

for any $\sigma(T_1, \ldots, T_K)$-measurable random variable C. The inequality follows because the sample median $\hat{m}$ minimises the sum of absolute errors away from the sample points $\{t_1, \ldots, t_k\}$. The proof is completed by picking $C = T_b^{(CF)}$. □